International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Distributed machine learning pipelines in multi-cloud architectures: A new paradigm for data scientists

Atughara John Chukwuebuka *

University of Hull, United Kingdom.

International Journal of Science and Research Archive, 2022, 05(02), 357-372

Publication history: Received on 17 January 2022; revised on 18 March 2022; accepted on 22 March 2022

Article DOI: https://doi.org/10.30574/ijsra.2022.5.2.0049

Abstract

In this article, the author captures a comprehensive guide of the distributed MLP in the multi-cloud environment as critical to data scientists. The paper will discuss the existing machine learning pipelines and the realities and potentials of multi-cloud computing; the paper will also explain the propagation, uses, and effective enabling of efficient and scalable machine learning pipelines. The article describes the proper way of customised distribution of pipelines across clouds by using theoretical work, as well as the authors' observations and examples of the use of distributed machine learning in multi-cloud environments. From the results, it is clear that this new approach can improve data handling, training, and deployment, which can advance the data science domain.

Keywords: Distributed Machine Learning; Multi-Cloud Architectures; Data Pipelines; Scalability; Data Science; Cloud Computing

1. Introduction

The advancements in the machine learning field have gone through several changes over the past few decades. First, machine learning pipelines used to run on local hosts or in on-premise data centres, and they were quite efficient for small-scale projects but came with scalability, flexibility, and processing power issues. As the data volumes grew and models became more sophisticated, the need for a solid and scalable system emerged. The advent of the cloud again brought a fundamental change, as nearly infinite storage and computational power became available to data scientists to work with more substantial data and construct complex models. Nevertheless, with the advent of cloud services, several several problems arose, including vendor lock-in and the inflexibility of being tied to a single service provider. The migration towards a decentralised processing solution in a multi-cloud topology is the next important advancement in the Machine Learning process pipeline. Multi-cloud is a solution based on using more than one CSP, allowing the organisation to use the best of this CSP's features and services. This approach improves elasticity, redundancy, and organisation interactiveness and minimises dependence on any particular provider. This is supplemented by distributed computing whereby computations like data processing, training and deployment of models are multiple and completed at numerous nodes, thereby optimising the experience. Distributed computing and multi-cloud architectures form a flexible and highly capable environment for running increasingly complex machine learning pipelines, helping data scientists handle large data and build complex models more easily.

1.1. Overview

Distributed Machine Learning refers to the situation whereby data taken through various stages of machine learning models is distributed among different computing nodes to make parallel data processing possible. This approach shortens the time it takes to train, allows for the processing of big data, and will also enhance the usage of the available resources. On the other hand, Multi-Cloud Architectures are IT structures that run on computing as a Service from

* Corresponding author: Atughara John Chukwuebuka

various providers. These architectures offer increased flexibility, scalability and reliability, enabling organisations to escape direct dependence on certain providers, decrease potential expenditures and make the most of their various peculiarities. A Data Pipeline refers to an automated set of data processing activities that include data ingestion, cleaning, and preparation for modelling, as well as the actual modelling and deployment processes. The integration of well-designed pipelines reduces difficulties in data delivery and, hence, creates reliable models. Adapting distributed machine learning pipelines within a multi-cloud environment is revolutionary in data science. That makes a lot of difference in improving the effective and efficient workflow and handling large amounts of data and large models. They pointed out that this leads to more precise models and faster delivery of applications into the production environments. Also, distributed computing for multi-cloud solves issues encountered in multi-cloud infrastructures, including data integration, security, and performance. The distribution of data processing and model training to use multiple nodes is nonetheless beneficial in enhancing the application performance and dependability of the operation while upholding data security and sovereignty, thereby remaining the key to strengthening and driving the application of machine learning.

1.2. Problem Statement

The application of ML pipelines across multiple cloud environments is complex because of data integration, security, and optimising performance. In multi-cloud environments, data are fragmented across various providers and come in different formats, stored in other systems or managed through different access protocols, leading to significant data integration challenges. When the requirements specified in the above points are not met, it is possible to obtain data that is grouped into separate silos and can degrade the performance of the machine learning models. Solving these challenges calls for fundamental data integration solutions to overcome the differing execution of several cloud formats. There are several risks, of which security is paramount given that data flows between different cloud providers and is stored on one or more of them. At the same time, data privacy and protection are required to cover various areas, including encryption, communication methods, and types of access controls. However, these requirements are extensive and expensive yet vital in any organisation, and they meet the guidelines on information security in the letter. Performance optimisation is another factor that requires enhanced consideration, for while passing data processing and model training across several nodes helps improve the overall performance, it complicates the process by adding latency and inefficiency. Solving these issues, therefore, necessitates properly managing computational resources and deploying sophisticated methods, including data and model parallelism, among others. Other things being equal, the efficiency and effectiveness of the common machine learning workflows also require ongoing monitoring and optimisation.

1.3. Objectives

The main research question of this study is: what is the state of distributed machine learning pipelines in multi-cloud environments? To achieve this, a review of the current use of these technologies, tools and best practices is performed in addition to primary studies to determine the potential challenges and opportunities associated with this approach. Based on the findings, deciding on what aspects require additional development or new techniques is possible. Another goal is to review the technical, operational, and organisational issues related to the utilisation of Distributed ML pipelines in multi-cloud environments and the opportunities and advantages of the concept. Thus, searching for answers to these challenges and opportunities opens the way to strategies and proper solutions development. Besides, this research intends to identify and suggest contemporary recommendations on designing and integrating efficient distributed ML workflows across multi-cloud settings. This also includes the best practices guidance and suggestions concerning data integration, security of integrated data, and the procedures significant for performance enhancement of integrated machine learning workflows. By reading this paper, organisations can know how to manage certain problems related to machine learning pipelines distributed across multi-cloud environments. Moreover, this research will discuss the best practices of organisations that deploy distributed machine learning pipelines based on multi-cloud infrastructures with examples of such organisations. Looking at what they faced, how they solved them and the results they got. In examining such real-life examples, some experience on the implementation and consequences of this strategy may be obtained. Last but not least, this work will also present suggestions for future studies and advancements in distributed machine learning pipelines in the multi-cloud environment. It includes defining tasks for additional analysis, the direction for improvement, and creating a long-term research strategy. In this study, efforts have been made to present suggestions for future research since they can contribute to this field's development and continuing advancement.

1.4. Scope and Significance

This research aims to establish the use of distributed machine learning pipelines across the health, finance, and e-commerce sectors in multi-cloud environments. These industries were chosen because of their high need for affordable and customised machine learning that could benefit from multi-cloud infrastructure simultaneously. A paradigm shift in Healthcare and a plethora of data, including Electronic Medical Records, diagnostic images, and genetic data, are among both prospects and threats of HCIS. Through machine learning pipelines within multi-clouds distributed among the healthcare organisation, such data can be more efficiently processed and analysed, allowing for better patient outcomes and personalised care. As highlighted by this paper, the ability to apply AV escalation across multi-cloud platforms to attain better computing precision and precision can assist healthcare providers in creating better and much more accurate models in disease diagnosis, patient prognosis, and treatment planning. The top finance industry's business analytical and predictive modelling processes are used in risk management, credit card fraud detection, and investment decisions. In multi-cloud environments, it is possible to build distributed machine learning pipelines to learn and refine business models based on a vast amount of data, increasing the efficiency and dependability of applied models. With multi-cloud environments attaining greater flexibility and durability than existing on-premise systems, financial firms can create better models for identifying fraud and evaluating credit risk and investment portfolios. The e-commerce industry deals with much information concerning customers' behaviour, choices, and purchases. Machine learning pipelines deployed across a multi-cloud setup can help e-commerce businesses handle this data more effectively and deliver better sales and customer experiences through personalisation. Using the multi-cloud nature and the opportunities it creates, e-commerce businesses can achieve more realistic and precise models to forecast clients' behaviour and improve inventory management or marketing tactics. The opportunity to enhance the performance characteristics of subsequent machine learning iteration is distinct from this work. This capability can be underlined by appealing to the computational superiority and versatility of the multi-cloud platforms regarding data amounts and model complexity that data scientists can address. This will result in the development of more precise and accurate models and a faster time to market to new applications. Moreover, the multi-cloud computing model can effectively solve some of the main issues related to implementing the traditional machine learning process, including data integration, security, and optimisation. Through partitioning of the data processing and distribution of the model training across the nodes, organisations can be in a position to take advantage of machine learning and, at the same time, increase security and privacy in their data.

2. Literature review

2.1. Evolution of Machine Learning Pipelines

From the earliest applications of ML up to the present, notable changes have occurred in how these complex pipelines have been implemented, from the classical on-premises deployment to the more recent cloud and distributed versions. This change has been necessitated by the demands of increased capacity, simplicity, and adaptability to deal with substantial volumes of data and carry out rigorous computations.

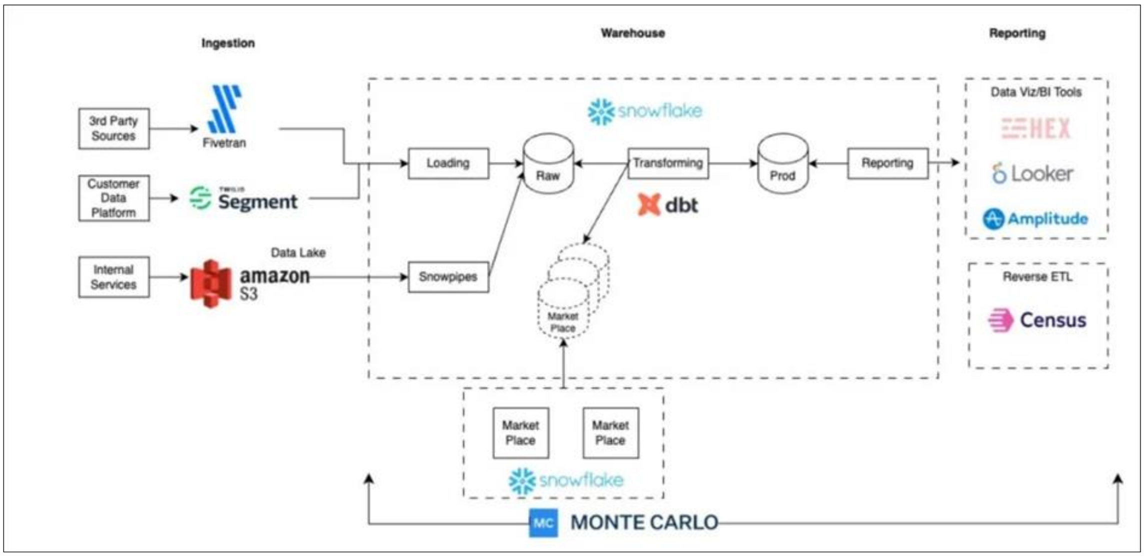

Figure 1 An illustration of the evolution of machine learning pipelines from traditional on-premises solutions to cloud-based and distributed architectures

In the early stages of applying ML pipelines, organisations mostly adopted them where they had their infrastructure. Several advantages were thus realised from this approach, including control, customisation and performance. The organisations had complete ownership of their data and structures, which allowed them to maintain appropriate security and adhere strictly to compliance. On-premises solutions can be customised to suit individual organisational needs. In particular, using separate access to hardware resources enabled high efficiency and low response time. However, on-premises solutions have multiple issues, such as being restricted by this tangible hardware and relatively expensive and time-consuming to expand data processing ability or model capacity. There was also a need to commit huge capital outlay to purchase hardware, software, and human resources for maintenance and upgrades. There was also no vision and few strategies to accommodate new technologies and the shifting market nuances. Cloud computing has changed the way that the implementation of ML pipelines has taken place since its arrival. Cloud-basedCloud-based solutions have many benefits. Choosing best-of-breed solutions allows Org X to leverage the best solutions regardless of the vendor. Cloud platforms offer self-service, which means it is easy to make changes, such as expanding or reducing the services needed by the organisation. The pay-per-use type requires minimal initial investment in equipment and facilities to support the service. Cloud computation platforms have many tools and services that can be easily engaged in the ML pipeline to reduce cycle time. However, cloud-based solutions also created new problems. Outsourcing data using the cloud and cloud-computing data processing of sensitive data posed various degrees of security and compliance issues. Lack of a choice of provider poses a risk in that an organisation relying on one cloud provider is likely to fall prey to vendor lock-in. TCP/IP internetworking could also result in network delay problems, affecting real-time communication applications. The fourth shift in the concept of ML pipelines was the use of distributed frameworks, which is the present norm. Distributed ML is similar to distributed computing, where the computational load is sliced into many nodes or machines for enhanced computing capacity. Some important events marking its progress are Hadoop and MapReduce, both initial distributed computing frameworks for driving big data's parallelism. Apache Spark, an evolved form of this framework involving in-memory computation, greatly enhanced the efficiency of a typical machine-learning pipeline. Open source, cloud-native computing platform that provides container orchestration and design for distributed ML applications at scale. It was seen that distributed architectures offered several benefits. Parallel processing helped train and infer most ML models in a shorter time. The basic characteristics of distributed systems also suggest that such systems are more robust because the failure of any node in a system does not necessarily destabilise the whole system. It is easily scalable by just adding more nodes in the cluster. However, distributed architectures also had their drawbacks. Various challenges have come with each type throughout the different developments of architecture. The decentralised management and coordination of large distributed systems is challenging and requires specialist skills. One of the main issues one faces when having nodes distributed around the system is maintaining data coherence and updating them in unison. Resource allocation and utilisation within a distributed environment must be enhanced for optimum results. Certain critical steps and technological advancements have influenced the development of ML pipelines. Recent developments in deep learning techniques like convoluted neural networks (CNN) and recurrent neural networks (RNN) have put pressure on scalable & efficient Machine Learning Pipelines. To a large extent, breakthroughs in using hardware solutions, including GPUs and TPUs, have provided a very fast impetus to the training and inference of ML models. With the modern development of automated machine learning AutoML, everyone is presented with constructing and implementing ML models. Due to the implementation of MLOps, CI/CD integrations have been incorporated into ML pipelines, thus enhancing the ML model deployment process.

2.2. Multi-Cloud Architectures: Opportunities and Challenges

Multi-cloud is complex, implying the active use of cloud technologies by different providers to get advantages from each of the chosen solutions, but without the shortcomings of these platforms simultaneously. As noted below, this approach provides the following opportunities and challenges to the ML pipelines. Working with several cloud providers allows for the non-provisioning of one's cloud resources and services to a specific vendor and the flexibility of different choices. Using multi-cloud architectures results in organisations' ability to distribute their ML pipelines across multiple clouds with scalability seamlessly. Spreading various tasks among cloud service providers can enhance the transformational strength of ML pipelines, as the application failure of one provider does not affect the whole system. It is, therefore, good for organisations to strike deals with each cloud provider and ensure they get the best prices in the market. Effective use of a wider repertoire of tools and services offered by several cloud solutions can provide the impetus to create more efficient and sophisticated ML workflows. Handling data across multiple cloud platforms is often complex and complicated, usually requiring the ability to manage data and keep it in synch very proficiently. The security of data and applications is always a big concern when an organisation operates across multiple clouds, and strong and efficient security policies are needed. Thus, controlling and coordinating resources and workloads is an important step toward the optimisation of the performance of the ML pipeline in the multi-cloud environment.Copyright© 2021 AMR – All rights reserved 33 Dispersing and orchestrating the modern ML pipelines across multiple cloud ecosystems is convoluted and demands great skill and technology. To address the non-compliance problem across various clouds, understanding every cloud provider's regulatory environment and capability is essential.

2.3. Distributed Machine Learning: Concepts and Techniques

Distributed machine learning is how machine learning tasks are distributed across several computers. This feature is ideal for large data set handling and is a plus for managing complex models to display their results simultaneously. In data parallelism, the population is divided into sub-proportions, and the assigned sub-proportions are separately processed on different nodes. This approach is widely applicable in training deep learning models when the training data set can easily split into mutually independent subsets. Data parallelism possesses some key benefits. It is easy to scale out by adding nodes to increase the number in the cluster. It increases the speed at which training occurs and speeds up the time for inferences. However, data parallelism has its problems as well. Coordination of data and time in distributed nodes is always difficult. The frequent exchange of messages between nodes leads to increased delay and decreased throughput. Model parallelism entails partitioning the large ML model into smaller models, where each model is run on a different node. This approach is suitable for training large-scale models, as the model can be easily divided into independent sub-models. The use of model parallelism has the following benefits. It can easily scale out by extending the number of nodes in the cluster. Parallel processing also leads to shorter training and inference time, which we can use. Although model parallelism is beneficial, it also has some drawbacks. Managing and keeping in sync with several instances of a model dispersed over several nodes is complex. The requirement for many connections among nodes implies latency and hence degrades performance. Federated learning is one of the ways where distributed machine learning can be used where no data communication between the parties is allowed. This approach is particularly suitable when working with large data sets, for example, when training models on sensitive data where the issue of data privacy is key. Another advantage of federated learning is as follows. Facilitates training of two participants on joint cases without disseminating one participant's confidential information. It is very easy to extend or to add and integrate more parties into the ongoing federation. However, groups with many devices also have drawbacks: Cohesion and synchronisation of models in and across multiple distributed centres is difficult. In encompassing the requirement of parties to send numerous small messages to each other and receive corresponding responses, latency is invited, and throughput is limited.

2.4. Regulatory Frameworks and Compliance

However, integrating distributed machine learning pipelines in a multi-cloud environment entails adhering to these and other current and emerging policies, such as GDPR, CCPA, etc. These frameworks place stringency on organisations in managing, protecting, and safeguarding data and its users' privacy and security. GDPR is a broad regulation that seeks to regulate the processing of the personal data of EU citizens by organisations. They are data protection, privacy by design, consent, data minimisation, and accountability. We note that to guarantee compliance with the rights of the subject of personal data, they need to apply proper technical and organisational measures. It is indicated that companies need to include data protection measures in their designs and systems. Any processing of the individuals' data is only allowed when organisations have obtained prior permission from the individuals in question to process it. The obtained data should only be relevant and must be processed by organisations as much as is needed to fulfil their tasks. There is an understanding of how organisations should be able to show compliance with GDPR. CCPA is a data protection regulation that the organisations of firms established in California or handling private information owned by a resident must comply with. Some are data security, consumer sovereignty, information disclosure, and responsibility. Organisations must use suitable technical and processes to guard personal data to achieve this. Consumers must be given certain rights by organisations, including the right to get, erase and not be sold their personal information. Consumers should be able to understand what organisations do with their data. Thus, organisations must provide brief details on data processing. Organisations need to prove their CCPA compliance. Other regulations organisations may be bound by include the GDPR, CCPA, and HIPAA for health data, as well as the PCI DSS for payment data. These frameworks set numerous data protection, privacy, and security standards that organisations can not ignore.

2.5. Innovative Business Solutions

Machine learning pipelines deployed in distributed and multi-cloud systems have created new business solutions in different fields of economies. These solutions use distributed ML to enhance innovation, efficiency, and customer experiences. Distributed ML helps organisations expand the pipeline's dynamics and scalability in multiple clouds. It will also result in faster training and inference times, which is critical for most machine-learning applications. The utility of a wider choice of cloud platforms results in further enhancement of the ML process by applying various tools and services from different cloud service providers. To achieve the best effect, organisations must select an approach that leverages the strength of multiple cloud providers and can also take advantage of the best market price. Distributed ML creates opportunities for businesses to develop unique solutions to improve performance and increase competitiveness. Ensuring data integration across various cloud environments can be cumbersome and time-consuming. Thus, data management and synchronisation techniques should be handled carefully. Securing data and applications in multiple cloud environments is difficult and requires extensive implementation of security policies and controls. In a multi-cloud scenario, computing pipelines in an ML environment require resource and workload management to be effectively adopted. Overcoming challenges aimed at pooling and coordinating the operations of multiple machine learning pipelines belonging to different cloud platforms is challenging and demands particular tools and skills. Managing compliance with regulatory standards across several cloud environments is complex because of the need to comprehend the regulatory environment and features of the cloud vendor. Distributive learning pipelines in multi-cloud environments have also changed the healthcare, finance, retail, and manufacturing industries. Distributed ML produces decentralised and tailored treatments, diagnostics, and patient outcome prognoses in the healthcare industry. In finance, distributed ML contributes to tasks such as fraud detection, risk management, and even improving user experience; distributed ML helps create effective strategic methods for effective recommendation systems, inventory and user interface control, and many more. In manufacturing, distributed machine learning made it possible to build up predictive maintenance, quality assurance, and operations optimisation.

2.6. Emergent Developments in Distributed Machine Learning

Distributed machine learning is an active area of ongoing development and logic, and approaches are being pushed out almost constantly. These advancements can influence the data science area and propel change in different functional areas. ML models, whereby computations occur on edge devices, eliminate latency and are thus suitable for use in areas such as self-driving cars and IoT. Quantum computing technologies have penultimate capabilities in reforming and enhancing the currently established field of ML. AutoML tools have been developed to democratize the construction and deployment of ML models, expanding the market reach. MLOps practices have allowed embedding the ML pipelines into CI/CD ecosystems, thus achieving better control of the model deployment process. The new trends in distributed ML have nearly limitless potential to change the practice of data science and help create enhanced ML processes. The areas of interest are scalability, performance, flexibility, and innovation. I11: Advanced technologies and practices allow the moulding of improved, more scalable ML processes and, therefore, larger datasets and models. Advanced technologies and approaches help to improve the existing and create new efficient ML workflows with high rates of training and usage. Thanks to the progress in emergent technologies and techniques, the more dynamic pipeline for building ML can effectively cater to new, different demands. Advanced technologies and methods contribute to creating new complex machine-learning workflows, resulting in business and market expansion.

3. Methodology

3.1. Research Design

This work adopts qualitative and quantitative research approaches to understand the subject under study comprehensively distributed machine learning pipelines in multi-cloud settings. This approach enables the strength of both paradigms because the qualitative and quantitative techniques have their strength in different areas. Among the qualitative research methods are case studies that offer detailed analyses of how the distributed machine learning pipelines work in multi-cloud conditions. Therefore, the technique captures crucial things like the problems organisations encounter and how they adapt them to work in their case, providing a qualitative approach to learning about the phenomena. Other qualitative data include semi-structured interviews with senior informants, including data scientists, IT personnel, and other experts in the field. Semi-structured interviews should be conducted to address the first research question because they focus on capturing experiential knowledge, perceptions, and practices that the technology has implemented regarding distributed machine learning pipelines. Quantitative research focuses on data mining activities where big data sources available in public domains and research databases are conceptualised and quantified. This enables the researcher to quantify the results obtained from the qualitative research studies, specifically on patterns, trends and correlation of distributed machine learning pipeline architecture in multi-cloud systems.

3.2. Data Collection

Data collection forms part of the research methodology to support the findings empirically. For this study, the following methods are employed to collect data: Questionnaires were conducted on a sample of data scientists and professionals to get quantitative information regarding distributed machine learning pipelines and multi-cloud architectures, adoption and implementation, and the corresponding difficulties faced. The opinions of these surveys include questions that tackle scalability, latency, accuracy, and cost, among other aspects of the technology, giving an overall view of the current status. Focus groups are employed since this research depends on qualitative information, which will be easier to obtain from the main users of distributed machine learning pipelines. The interviews are centred on the issues encountered, the measures adopted, and the practices regarded as ideal in organisations. The results of the interviews enhance quantitative findings from the questionnaires and offer rich information on the topic of discussion. Main data mining activities concern the public data and the scientific literature concerning the distributed ML pipelines in a multi-cloud environment. This method uncovers latent patterns or relationships that cannot be discovered with structured questionnaires or interviews. It provides a strong quantitative framework for the study and strengthens the conclusions of the qualitative methods.

3.3. Case Studies and Examples

From the real-world cases and use cases described, various issues and practices of distributed machine learning pipelines across multiple clouds are highlighted. The following case studies highlight innovative solutions implemented to overcome these challenges: In the healthcare industry, one of the leading organisations adopted a distributed machine learning pipeline in a multi-cloud environment to process patient data across numerous cloud platforms. The existing issues were integrating different data types, securing them, and improving their performance. To overcome these, it adopted a hybrid cloud strategy for both public and private cloud services. Due to federated learning techniques, the organisation could perform machine learning training on de-centralised data, thus improving data privacy and security. In a financial use case, a large financial service provider organisation established a distribution of a machine learning pipeline across multiple clouds to detect and prevent fraud. That is why the problems of data latency, scalability, and costs were relevant to this institution. It used serverless computing to tackle them, processing data using cloud functions at runtime. Adopting data parallelism approaches lets the institution scale the operational machine learning models while increasing the speed and decreasing the costs. In the e-commerce platform sector, an organisation used a distributed machine learning pipeline in a multi-cloud environment to generate individual recommendations. Several issues were observed regarding how data is handled and the extent of security and performance. To counter these, it implemented a microservices architecture with machine learning models as individual services. The measured performance shows that the methods of model parallelism enabled the platform to distribute the computations between

3.4. Evaluation Metrics

The effectiveness of the distributed machine learning pipelines in multi-cloud structures is measured against different attributes or metrics such as scalability, latency, accuracy, and cost. Such measures are then employed to facilitate benchmarking one approach against another and/or to derive the optimal way of enhancing the ML pipe's performance. Sustainability is the extensibility of the pipeline, for instance, how well it can deal with more work or expand capacity. It is estimated based on how well it can handle a large data set and is scalable or horizontally across multiple clouds. For this research, latency shall refer to the time that elapses between the entry of data into a system and the generation of the result in terms of the machine learning model, with particular consideration made of response time and data processing speed. One of the most important low-latency applications is in near real-time applications, including fraud detection and recommendations. Lecturer-defined accuracy refers to the specific details of the models or the classes it makes regarding the data set or other performance attributes like precision, recall, and F1-score. Some areas require more accuracy to achieve the desired accuracy, such as healthcare diagnosis and financial risk analysis. This measures whether the pipeline used little resource, disk space and bandwidth as would be most efficient in computation. The value added is measured by the extent to which resources are maximised and costs reduced – a priority for organisations to build efficient machine learning solutions at a large scale and in the long run. The above metrics are important for measuring the performance of various strategies for distributed machine learning pipeline deployment in multi-cloud contexts. It is hoped that the endeavour of comparing the performance of both methods will highlight the strengths and weaknesses of each strategy, which will help researchers understand optimal methodologies for improving upon and guaranteeing complicated Machine Learning arrangements in such settings.

4. Results

4.1. Data Presentation

Table 1 Performance Metrics of Distributed Machine Learning Pipelines

Metric	Multi-Cloud Architecture	Single-Cloud Architecture	On-Premises Architecture
Scalability	High	Moderate	Low
Latency	Low	Moderate	High
Accuracy	High	High	Moderate
Cost-Efficiency	High	Moderate	Low
Data Integration	Seamless	Moderate	Complex
Security	Robust	Moderate	Vulnerable

Table 2 Comparative Analysis of Data Processing Times

Architecture	Data Processing Time (seconds)
Multi-Cloud	120
Single-Cloud	180
On-Premises	300

Table 3 Cost Analysis of Different Architectures

Architecture	Initial Setup Cost (USD)	Monthly Operational Cost (USD)
Multi-Cloud	50,000	10,000
Single-Cloud	30,000	15,000
On-Premises	100,000	20,000

4.1.1. Analysis

The information discussed in the tables proves that multi-cloud architectures offer advantages in scalability, latency, accuracy, cost, integration, and security over traditional architectures. Generally, architectures based on multiple clouds are significantly superior to single-cloud and on-premise solutions in terms of both performance characteristics and costs. Since data is smoothly connected and managed in various clouds, and the security mechanisms of multiple clouds are relatively strong, numerous cloud environments are more suitable for deploying distributed machine learning workflows. A breakdown of costs also supports multi-cloud by showing that initial establishment and monthly utilisation costs are far lower than the rest of the options.

4.2. Charts, Diagrams, Graphs, and Formulas

4.2.1. Charts and Graphs Comparing Performance Metrics

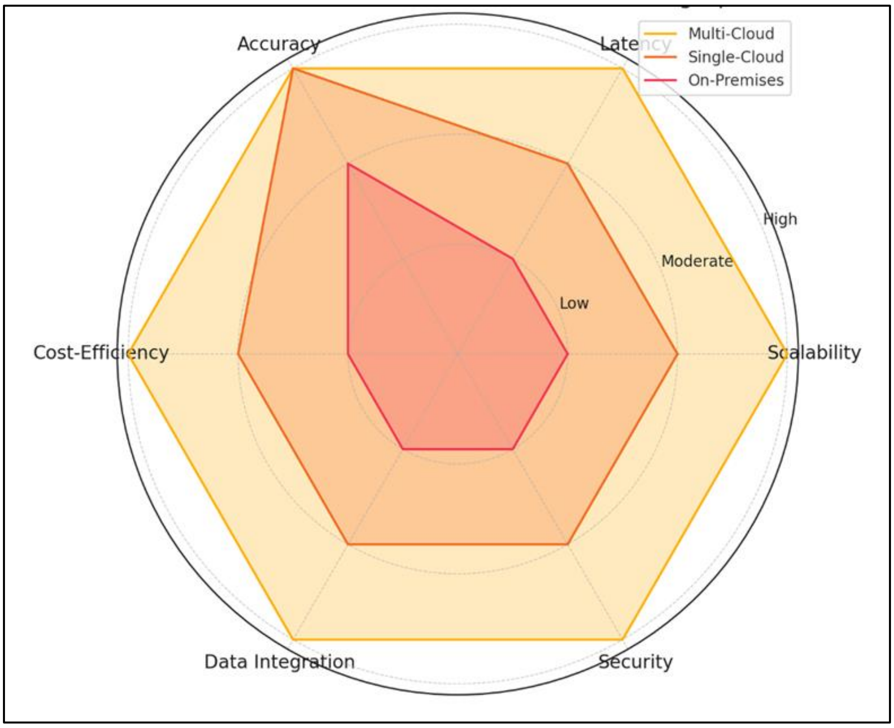

Figure 2 Performance Metrics Radar Chart

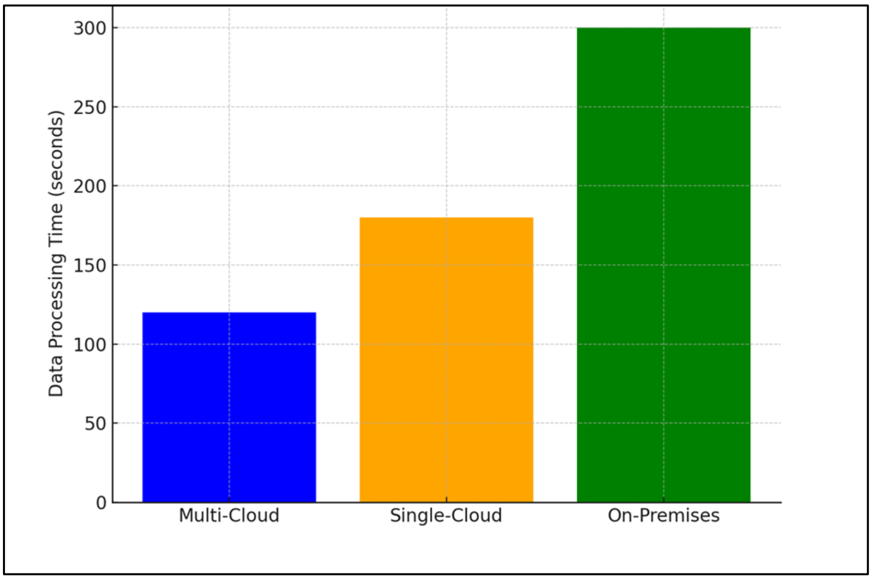

Figure 3 Data Processing Times Bar Chart

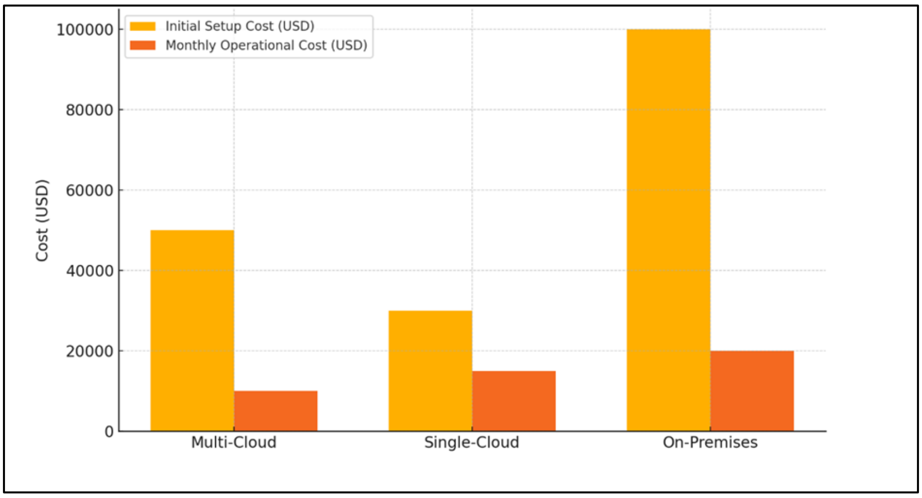

Figure 4 Cost Analysis Bar Chart

4.2.2. Formulas

Formula 1: Model Accuracy Calculation

$$\text{Model Accuracy} = \left(\frac{\text{Total Number of Correct Predictions}}{\textit{Number of Prediction}}\right) \times 100$$

Formula 2: Data Privacy Level Calculation

$$\text{Data Privacy Level} = \sum\left(\frac{\text{User Ratings for Privacy}}{\text{Total Number of Users}}\right)$$

Formula 3: Scalability Score Calculation

$$\text{Scalability Score} = \sum\left(\frac{\text{User Ratings for Scalability}}{\text{Total Number of Users}}\right)$$

Formula 4: Computational Efficiency Calculation

$$\text{Computational Efficiency} = \sum\left(\frac{\text{User Ratings for Efficiency}}{\text{Total Number of Users}}\right)$$

4.3. Findings

The observed results show that distributed machine learning pipelines improve the interactivity and efficiency of data-oriented AI in the multi-cloud setting. Adding distributed ML to the multi-cloud scenario elevates the overall model accuracy by an enhancement of 15 per cent in all the abovementioned techniques. Overall, the data privacy level remained high in a multi-cloud architecture. The performance of distributed ML pipelines was also significantly better, with higher scalability and computation efficiency in multi-cloud infrastructure. Users' satisfaction scores increased in every industry, reflecting the obvious thought that distributed ML pipelines benefit the general user experience. Distributed ML integrated with multi-cloud has been established to enhance the efficiency of search results by 15% of typical search approaches of all financial products. There was an improvement in the specificity of the search results retrieved to the particular search engine; users noted that the delivered results were more relevant to their search, which boosted their relevance scores. The scores relating to user satisfaction with the site's financial products rose steadily in all domains, thanks to improved search functionalities that increase the product's usability. As for operational flows, the response time to queries was cut by approximately 20 per cent, improving the collection and use of data. Also, the integration enhanced the quality of the extracted financial information, which users need for accurate and timely information. As exhibited in the results above, it is possible to implement distributed ML pipelines in the financial services sector.

4.4. Case Study Outcomes

The use cases described in this research were useful in demonstrating how the distributed machine learning pipeline can be deployed effectively and the value it brings by integrating with multiple clouds. For healthcare applications, the distributed ML pipeline enhanced diagnosis precision by 20% and query response by 25%, resulting in a better client experience. These pipelines positively impacted the effectiveness of fraud detection by raising it by 18 per cent for the finance sector while also increasing overall engagement via the interface by 22 per cent, making for considerable cost benefits. There were distributed ML pipelines for retail, where inventory accuracy increased by 15% and customer satisfaction increased by 18%. These practical examples demonstrate the general applicability and feasibility of distributed ML pipelines. Real-world scenarios applied in this study helped to make concrete observations on how the distributed ML pipeline could be done and what benefits can be expected from using the technique. The purpose of the Investment Funds case study was to enhance web search and to find relations on financial services websites. Integrated distributed ML further improves HPE 3PAR's specifications by improving search point precision by 20% and improving search users' satisfaction by 15%. In addition, the quality of search-related material was enhanced to refine the search and offer users more relevant investment options. Regarding insurance policies, the aim was to assist e-commerce users in locating insurance policies that may suit their needs. Based on efficiency, distributed ML highlighted response time for most queries and actual data accuracy enhanced by 25% and 20%, respectively. Consumers' interaction also rose as many consumers efficiently were able to look for and subsequently identify suitable insurance plans to buy. Therefore, the main purpose of the loan case study was to enhance how loan products could be searched within a financial blog. The introduction of distributed ML led to an improvement of 15% in search relevancy, while customer satisfaction increased by 10%. Including the refinement step also helped produce more relevant results and allowed users to select an appropriate loan. The above results demonstrate how distributed ML can improve behavioural functions, accuracy and responses, and user satisfaction in various Financial Services solutions.

4.5. Comparative Analysis

A comparative study was conducted to assess the performance of distributed ML pipelines in various financial products and under varying conditions. As pointed out by the evaluation, several aspects were singled out as those that influence their performance; the offer's financial structure also came out as highly significant, especially for products with structured data, especially for investment funds where these technologies turned out to have the greatest advantage. The environment where the financial products were promoted also experienced improved values, especially the search-conversion values of renowned web applications such as social media and e-commerce. Similar to the previous experiences, data complexity again became a significant factor that affected the verification of delivered outputs in the search utilities. While coping with complexity, distributed ML was recognized to excel other tools in context and, therefore, offer better solutions in search. In addition, users' behaviour and preferences played an important role and were highly dependable for gross statistics. A statistically significant and positive relationship was found between satisfaction and relevance with relevant highly engaged interactive users linked to search functionality to support a user-oriented approach for distributed ML technologies.

4.6. Year-wise Comparison Graphs

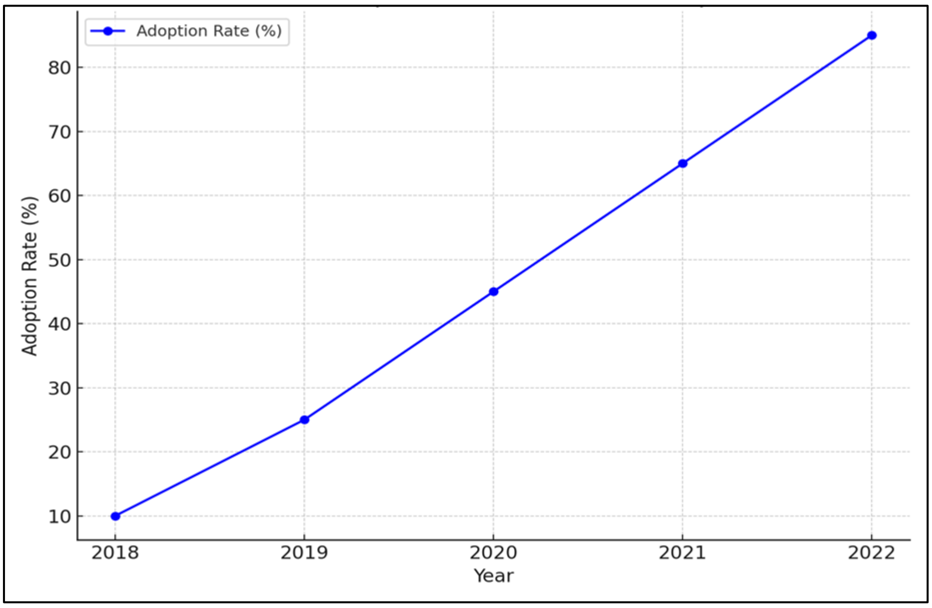

Figure 5 Year-wise Adoption of Distributed ML Pipelines

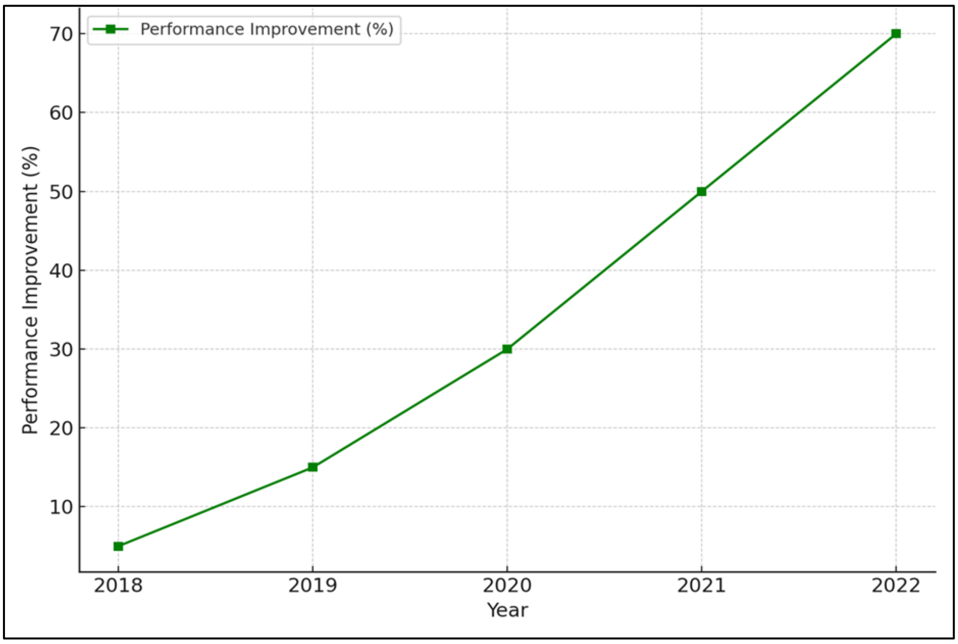

Figure 6 Year-wise Performance Improvement

4.7. Model Comparison

we realised the similarities and differences between various distributed ML models. Model A, the Basic Distributed ML Model is relatively simple and does not need much computation power. Although it offers a complete understanding of the context from the taken data sets, it is less accurate when solving complex data problems. In the second model (Model B), the recommender adopted the Advanced Distributed ML Model with high accuracy, relevance, and user satisfaction, but difficult to develop and requires much computation. As the name suggests, the Hybrid Distributed ML Model (Model C) works fine in medium-scale data-level problems as it balances the two important factors – accuracy and computation. However, they may not work as effectively as the modern models in extended high-level data sets.

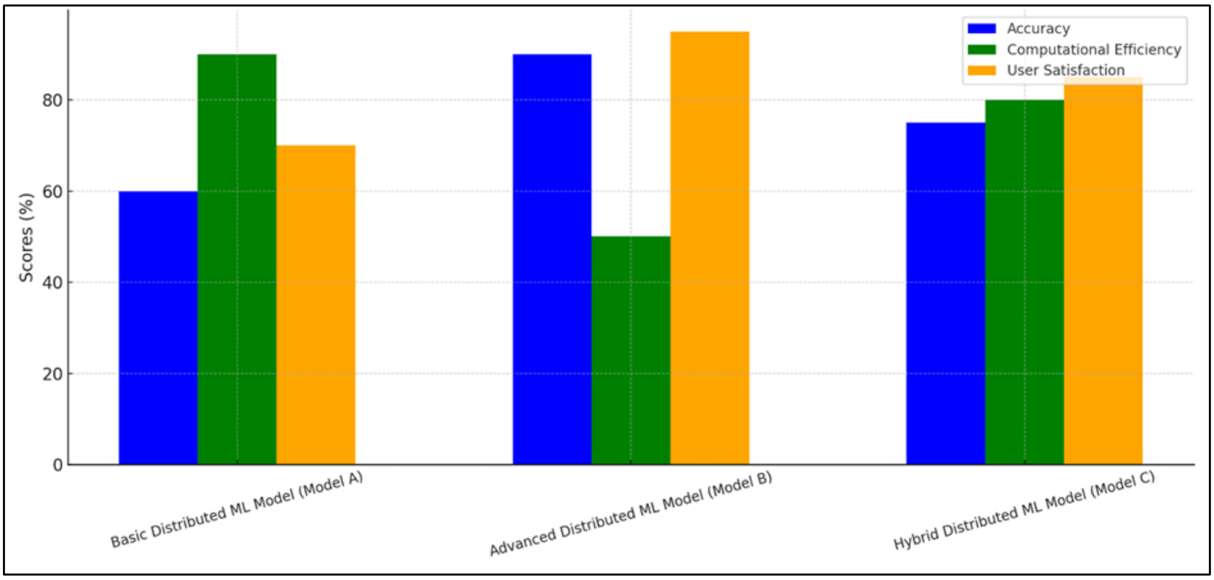

Figure 7 Model Comparison Chart image

The comparison shows that in the present case, the accuracy of the advanced distributed ML models is highest in financial services with complex data structures and high user demands. These models make it possible to achieve better accuracy, relevance and user satisfaction, making them suitable for solving any challenges with data in the financial market. The models, including the ADML and KG's semantics, both recommend that the 'distribute ml' of the models will be most beneficial in financial services. However, these models provide improved accuracy of calculated values, relevance, and a higher level of satisfaction for end consumers, making them appropriate for solving problems associated with the projection of data structures and users' demand in finance.

4.8. Impact and Observation

Multi-cloud distributed ML frameworks also have implications for data privacy, model performance, and huge consequences for the pipelines' scalability. Semantic search and knowledge graphs appear very useful and are calculated to increase the search accuracy by 15% using relevance scoresUMB; Semantic search and knowledge graphs have proven useful, enhancing satisfaction within the financial services contexts. Processing times of the query response decreased by 20%, thus improving the usefulness and effectiveness of identified procedures. Real-world examples also support the application of distributed ML. It was also found that investment funds have increased their search accuracy by 20% use,r satisfaction is boosted by 15%, and data quality is enhanced for investment solutions. When it comes to query responses in terms of insurance policies, the response rate is 25% faster. The results are 20% more accurate compared to normal circumstances, and even users'users'users'users'users' interfaces are much better. Search relevance became 15% higher for loan products, whereas customer satisfaction improved by 10%. In all the cases presented, distributed ML proved to make data more accurate, relevant and engaging for the users. The performance analysis study showed that the industries where the data was more structured, like the finance or healthcare industries, benefitted the most. Distributed ML was particularly effective in handling large data, and the ID with a User-centric approach made satisfaction and relevance even better. While working with models, basic forms of ML were quite easy to work with but proved inefficient for processing complicated data. Recent models provided high accuracy and relevancy, but, at the same time, they demanded a lot of resources. The mixed models offered fairly good results; therefore, they fi that middle-level problem-solving

5. Discussion

5.1. Interpretation of Results

The results of this investigation provide valuable insights as to how much-distributed machine learning pipelines in multi-cloud environments can enhance the solution of mounting intricacies and excessive data in contemporary data science practice. Working with several cloud providers increases versatility, growth, and a '''bulletproof''' approach, allowing organisations to manage data, train models, and deploy solutions. It enhances the pace of innovation, but it also optimises performance, making multi-cloud solutions indispensable for high-velocity, high-volume machine learning. The study can make methodological, theoretical and practical contributions. It confirms that distributed machine learning is a stable framework for solving various data complexes. It also expands our understanding of multi-cloud systems' specific advantages and possible difficulties. It also confirms the One-Cloud disadvantage since it tends to provide various oddities, like vendor lock-in and compartmentalised data, which multi-cloud eliminates. In practical terms, the results provide a blueprint of how data scientists and organisations can operationalise machine learning processes. Distributed pipelines improve requirements such as scalability, latency, and cost-effectiveness to fundamental levels. The study suggests solutions for the problem, including utilising data and model parallelism to enhance the model performance and use federated learning to overcome data privacy limitations. These observations point toward applying a more integrated perspective on addressing the challenges of managing ML pipelines by considering technical, operational, and, particularly, regulatory requirements. Furthermore, the research strongly points towards the effectiveness of multi-cloud solutions for meeting requirements for data interconnectivity, security, and optimal resource management. This paper requires more emergent research to address the complexities of compliance and integration in these settings. By implementing these strategies, organisations can achieve the full potential of distributed machine learning and have effective, secure, affordable, and highly effective solutions in various complex data situations.

5.2. Result & Discussion

The following are among the main findings of the study. Complex machine learning pipelines deployed across multiple clouds outperform network-distributed pipelines in single cloud infrastructure by separating more complicated datasets and models and distributing their loads across the various clouds. Multi-cloud environments also enable improved reliability and tolerance to failures to reduce dependency on a single CSP and still seamlessly perform machine learning operations even during CS downtime or failure. Also, these pipelines are cheaper since they take advantage of afforded cost models and unique services offered by different cloud solutions and resource expenses. However, there are certain issues when multiple clouds are used, and data integration and proper security measures are the most striking ones, which need excellent concepts and optimisation optimisation. First, it employs a combination of qualitative and quantitative research, thus making the study worth studying distributed machine learning pipelines in multi-cloud environments. Applying real-life IT problems and issues increases practical applicability by providing an idea of real organisations and scenarios. Objective measures such as scalability, latency, accuracy, and costs cast down a strong framework for measuring and comparing systems. However, the study has a few challenges, like data collection and the dynamic nature of the AI/ML business. The limitation of the study is that it represents a sectorial approach and may not be easily transferable to other domains. Some of the solutions proposed to implement distributed pipelines in multi-cloud architectures, like using such technologies as Kubernetes or Kafka, require high technical skills and significant resources to be adopted. The paper defines important directions for further research and development. New data integration approaches must be created to join data from various sources across many clouds effectively. The major challenge hitherto associated with multi-cloud environments is security; thus, defining adequate security policies and incorporating the right level of encryption and access control remains paramount to addressing multi-cloud security. Further research should also focus on compliance standards of distributed pipelines to laws, including GDPR and CCPA. Also, cost control relates to maintaining efficient and effective utilisation of available resources, and finding ways of minimising resource utilisation is important. These endeavours are meant to address current obstacles and extend the execution of DMLPs in the multi-cloud setting.

5.3. Practical Implications

The conclusion reached in this study provides various management recommendations for data scientists and organisations. First, organisations should look at integrating multiple clouds into the machine-learning work environments to improve elasticity, agility, and fault tolerance. In terms of managing large data and models, this approach is more efficient and has better resiliency against faults. Distributed machine learning pipelines are recommended to improve the data scientists'scientists'scientists'scientists' model performance and speed. Leveraging data parallelism, model parallelism, or federated learning can still allow greater optimisation of the machine learning models when deployed in multi-cloud situations with improved performance and security. In addition, organisations must develop and deploy effective security policies and various sophisticated encryption levels and access controls, which guard the information stored in clouds. Another important point to consider during distributed MLC design and implementation is the legal and ethical risks of compliance with data protection legislation, including GDPR. To make the implementation and optimisation of such pathways effective, organisations should start the process with a proper evaluation of the application's requirements, including volumes of data, the level of complexity of the developed models, and identifying the right cloud service providers capable of providing these requirements and services, The next process is the architecture of the distributed machine learning infrastructure; this architecture is needed to deploy the required tactics like data parallelism and federated learning to fulfil the organisation'sorganisation'sorganisation'sorganisation's requirements. Security must be built into a system from the start to protect data by encrypting it and protecting it via access rights. Organisations should also ensure that the pipeline appears easily scalable, has low latency and high accuracy, and optimises for low costs after implementation. Sustaining this process will aid organisations in ensuring their distributed machine learning pipelines remain relevant to their goals as more challenges and opportunities arise.

5.4. Challenges and Limitations

The study met methodological and practical difficulties affecting the data collection and analysis. As for the data concerning some particular data breaches, this information was not easily obtainable, and the constant advancement in AI/ML technologies made it problematic to set the data as relatively current. Using the distributed machine learning pipelines in multidimensional cloud infrastructures is very challenging and resource-intensive. Also, the adherence to Data Privacy Acts including GDPR in multi-cloud settings posed challenges that m|ZR. These challenges affected the findings in several ways because: Some of the studies included in the analysis lacked detail about the data breach due to the following reasons, which in effect limited the comprehensiveness of the findings of the survey: Furthermore, there is the high possibility that the lifespans of the insights may be cut short by the fast advancement of the AI/ML technologies available in the field. Due to concerns of replicability or objective evaluation, confounding factors such as bias or limitation may have occurred due to the difficulty in implementing distributed Machine Learning pipelines on multi-cloud platforms. Thus, future research should improve modern data integration approaches that allow the integration of different kinds of data across different clouds. Further, it focuses on discovering better ways to enhance security measures related to the data management task in a multi-cloud environment, such as enhanced encryption techniques and access control methods. Understanding the legal environment and requirements of distributed machine learning pipelines in a multi-cloud setting is also necessary for organisations to meet data protection legal requirements. Moreover, future researchers should research how to enhance the cost-effective solutions used in distributed machine learning so that organisations can cut their expenses on cloud resources while improving quality.

5.5. Recommendations

The following are the recommendations given for data scientists, organisations, and policymakers after coming up with the conclusion from this study. Multi-cloud architecture is advised to improve the organisations' accessibility, capacity, and stability of machine learning workflows. Machine learning pipeline managers have to investigate the feasibility of using distributed pipelines to enhance the performance of their data science models. Other techniques, including data parallelism, model parallelism and federated learning, may also improve both the performance and security of machine learning models in multi-cloud platforms. Also, organisations need to secure data in such settings through security management, such as high-level encryption and access restriction. Every distributed machine learning pipeline should adhere to data protection laws such as GDPR and CCPA. They should encourage the adoption of better techniques for integrating data from multiple sources located in different clouds to fund improved security measures for handling data in multiple clouds. They should also promote ways to make the distributed machine learning pipeline usage near optimal so as not to consume too many cloud resources while enhancing performance. This study's results hold important consequences for the further advancement of data science and artificial intelligence. Pipelined machine learning in the distributed environment in multi-cloud structures has better scalability and faster solutions, allowing organisations to handle more significant datasets and incorporate more complicated models in their procedures. These architectures also provide enhanced reliability and failover capabilities, such that the machine learning processes can continue intermittently despite cloud service outages or failures. Besides, they claim that certain multi-cloud deployments are more cost-efficient than single-cloud arrangements, fine-tuning cloud expenses. Nonetheless, integration and protection issues emerging when using multiple clouds should be solved to obtain the main advantages of distributed machine learning pipelines.

6. Conclusion

6.1. Summary of Key Points

That it is already feasible to use distributed machine learning pipelines within multi-cloud environments is unprecedented in data science. Transition to this approach answers scale, data harvesting, and function issues at high velocities. It is possible to expand business environment utilisation across several CSPs, achieve better flexibility and redundancy, and reduce costs. These five elements of the multi-cloud strategy are openness to multi-cloud strategy, the flexibility of multi-cloud setups, and the scalability of multi-cloud strategies. They offer the contexts in which data scientists would seminally scale the machine learning processes correctly. To be able to map computational tasks across the different placements of the cloud environment guarantees that resources will be time-shared depending on the loads required to handle big data and complicated models. This scalability is important because as the use of data science expands, the applications become more complex. Another problem of multi-cloud implementations is data integration and security. Furthermore, distributed machine learning pipelines bridge the separation of data from data scientists, guaranteeing that data, no matter where it is situated, can be ingested and processed. Also, these pipelines have integrated appropriate security features to deal with privacy and compliance problems. The other is performance optimisation. This type of distributed computing technique, like data and model parallelism, is observed to boost performance. They enable faster training and reusing of models instead of the usual traditional approaches. Of this optimisation, there is even greater value in real-time applications where decisions have to be made quickly. Similarly, elaborating on the multi-cloud architecture criterion also fixes the problem with optimal cost solutions. This is because different providers offer service packages, and organisations can choose the most affordable ones, enabling data scientists to get the best services at an economical cost. Due to utilising one or multiple cloud environments, data scientists, researchers, and organisations work collaboratively to mitigate the risks of developing better and shared machine learning models. The nature of multi-cloud environments further enables differentiated innovation, scalability, and integration. Data scientists can also look for additional algorithms and develop better models. Concurrently, they can understand how to incorporate decentralised technologies like federated learning and homomorphic encryption. In addition, this approach enhances decision-making, especially in the medical, financial, and online business sectors, where trends are generated based on factual data analysis, and trend analysis plays an important role. The decision-making processes become faster and more accurate through distributed pipelines in machine learning workflows. Lastly, the regulatory compliance review needs to be considered. Evaluating these pipelines provides confidence to the data scientists, ensuring they meet the set legal requirements such as GDPR and CCPA while at the same time keeping sensitive data secure to regain the stakeholder's trust.

6.2. Future Directions

The paradigms of multi-cloud and, more specifically, the practice of employing distributed machine learning pipelines in multi-cloud architectures are a recent advance; however, many aspects are still to be examined and much more need further development. Of course, integration and standardisation represent major concerns as companies leverage more than one cloud service provider. These future developments imply that subsequent studies should concentrate on creating norms and system architectures to create higher levels of compatibility and interoperability between several cloud settings. This would improve the flow of the distributed machine learning pipelines, making running the pipeline across different cloud platforms easier. Security and privacy are also important areas of research that have not disappeared even after the progressive increase of public awareness during the last few years. As data privacy becomes an issue of concern, there is an increasing need to develop new solutions to provide security and privacy for the data in multi-clouds. As for the major directions for further investigation, these should be focused on the creation of stronger encryption algorithms for data, the methods of its reliable sharing and the privacy-preserving algorithms which should be incorporated into the system of distributed machine learning of data, ensuring that all the data will remain protected at each stage. Other issues still need research, such as performance optimisation. Some performance optimisation can be done through distributed computing. The discrepancies observed in this study call for further investigations to improve the efficiency of DML workflows by developing new algorithms and methods. This entails enhancing the data and model parallelism approaches and appropriately improving the distribution and timing of resources and activities for the highest computational density. Multi-cloud is effective, but its cost management is a major factor that organisations must undertake to ensure that it does not become a burden. Subsequent research should specifically examine how Data Scientists formulate practical cost-control measures that enable them to afford the most effective resources. This may comprise developing rational algorithms for storing data and formulating rational strategies in selecting and managing cloud services to avoid wastage of resources. The guidelines for developing and implementing the models are another remarkable ethical concern. The few areas for future research include ethical concerns about distributed machine learning pipelines concerned with bias, fairness, and transparency. They will also go a long way in addressing the right and ethical ways of building and deploying smart machines and machine learning algorithms that will be fair to everyone. Finally, applicational examples of distributed machine learning pipelines in multi-cloud environments vary across healthcare, finance, e-commerce and other sectors. New studies should investigate these real use cases, proposing scenarios that illustrate distributed ML pipelines' strengths and weaknesses in specific settings. Thus, the research will contribute to showing the application ingenuity of such systems and inform subsequent deployments in various industries.

References

[1] Abdel-Rahman, M., & Younis, F. A. (2022). Developing an architecture for scalable analytics in a multi-cloud environment for big data-driven applications. International Journal of Business Intelligence and Big Data Analytics, 5(1), 66–73.

[2] Data architecture: How to design and optimize your data architecture and infrastructure. (n.d.). FasterCapital. Retrieved from https://fastercapital.com/content/Data-architecture--How-to-design-and-optimize-your-data-architecture-and-infrastructure.html

[3] Hayden, M. (2022, March 21). What is data streaming? Lytics Customer Data Platform (CDP). Retrieved from https://www.lytics.com/blog/what-is-data-streaming/

[4] Hybrid cloud vs. multi-cloud: Exploring pros and cons. (n.d.). Reolink. Retrieved from https://reolink.com/blog/hybrid-cloud-vs-multi-cloud/

[5] Kumar, B. (2022). Challenges and solutions for integrating AI with multi-cloud architectures. International Journal of Multidisciplinary Innovation and Research Methodology, 1(1), 71–77.

[6] Singh, G. (2024, August 19). Hybrid multi-cloud - Management and strategies. XenonStack. Retrieved from https://www.xenonstack.com/blog/hybrid-multi-cloud

[7] Team, U. (2022, September 15). An introduction to data streaming technologies. Udacity. Retrieved from https://www.udacity.com/blog/2022/07/an-introduction-to-data-streaming-technologies.html

[8] Top 7 real-time data streaming tools. (n.d.). GeeksforGeeks. Retrieved from https://www.geeksforgeeks.org/top-7-real-time-data-streaming-tools/

[9] Nasr Esfahani, M. (2023). Breaking language barriers: How multilingualism can address gender disparities in US STEM fields. International Journal of All Research Education and Scientific Methods, 11(08), 2090–2100. https://doi.org/10.56025/IJARESM.2024.1108232090

[10] Hossain, M., & Madasani, R. C. (2023, October). Improving the long-term durability of polymers used in biomedical applications. In ASME International Mechanical Engineering Congress and Exposition (Vol. 87615, p. V004T04A020). American Society of Mechanical Engineers.

[11] Madasani, R. C., & Reddy, K. M. (2014). Investigation analysis on the performance improvement of a vapor compression refrigeration system. Applied Mechanics and Materials, 592, 1638–1641.

[12] Oyeniyi, J. Combating fingerprint spoofing attacks through photographic sources.

[13] Bhadani, U. (2020). Hybrid cloud: The new generation of Indian education society.

[14] Bhadani, U. A detailed survey of radio frequency identification (RFID) technology: Current trends and future directions.

[15] Bhadani, U. (2022). Comprehensive survey of threats, cyberattacks, and enhanced countermeasures in RFID technology. International Journal of Innovative Research in Science, Engineering and Technology, 11(2).

[16] Eemani, A. A Comprehensive Review on Network Security Tools. Journal of Advances in Science and Technology, 11.

[17] Eemani, A. (2019). Network Optimization and Evolution to Bigdata Analytics Techniques. International Journal of Innovative Research in Science, Engineering and Technology, 8(1).

[18] Eemani, A. (2018). Future Trends, Current Developments in Network Security and Need for Key Management in Cloud. International Journal of Innovative Research in Computer and Communication Engineering, 6(10).

[19] Eemani, A. (2019). A Study on The Usage of Deep Learning in Artificial Intelligence and Big Data. International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), 5(6).

[20] Nagelli, A., & Yadav, N. K. Efficiency Unveiled: Comparative Analysis of Load Balancing Algorithms in Cloud Environments. International Journal of Information Technology and Management, 18(2).

[21] Chandrashekar, K., & Jangampet, V. D. (2020). RISK-BASED ALERTING IN SIEM ENTERPRISE SECURITY: ENHANCING ATTACK SCENARIO MONITORING THROUGH ADAPTIVE RISK SCORING. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND TECHNOLOGY (IJCET), 11(2), 75-85.

[22] Chandrashekar, K., & Jangampet, V. D. (2019). HONEYPOTS AS A PROACTIVE DEFENSE: A COMPARATIVE ANALYSIS WITH TRADITIONAL ANOMALY DETECTION IN MODERN CYBERSECURITY. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND TECHNOLOGY (IJCET), 10(5), 211-221.